Netflix Data Analysis And Visualization

```python
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import wordcloud
```

[11] `data=pd.read_csv('/content/netflix1.csv')`

`data.head()`

	show_id	type	title	director	country	date_added	release_year	rating	duration	listed_in
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	United States	9/25/2021	2020	PG-13	90 min	Documentaries
1	s3	TV Show	Ganglands	Julien Leclercq	France	9/24/2021	2021	TV-MA	1 Season	Crime TV Shows, International TV Shows, TV Act...
2	s6	TV Show	Midnight Mass	Mike Flanagan	United States	9/24/2021	2021	TV-MA	1 Season	TV Dramas, TV Horror, TV Mysteries
3	s14	Movie	Confessions of an Invisible Girl	Bruno Garotti	Brazil	9/22/2021	2021	TV-PG	91 min	Children & Family Movies, Comedies
4	s8	Movie	Sankofa	Haile Gerima	United States	9/24/2021	1993	TV-MA	125 min	Dramas, Independent Movies, International Movies

Next steps: [Generate code with `data`] [● View recommended plots] [New interactive sheet]

[13] `print(data.head(10))`

```
  show_id     type                             title            director  \
0      s1    Movie              Dick Johnson Is Dead     Kirsten Johnson
1      s3  TV Show                         Ganglands     Julien Leclercq
2      s6  TV Show                     Midnight Mass       Mike Flanagan
3     s14    Movie  Confessions of an Invisible Girl       Bruno Garotti
4      s8    Movie                           Sankofa        Haile Gerima
5      s9  TV Show         The Great British Baking Show    Andy Devonshire
6     s10    Movie                      The Starling     Theodore Melfi
7    s939    Movie       Motu Patlu in the Game of Zones     Suhas Kadav
8     s13    Movie                       Je Suis Karl  Christian Schwochow
9    s940    Movie              Motu Patlu in Wonderland       Suhas Kadav

          country date_added  release_year rating  duration  \
0   United States  9/25/2021          2020  PG-13     90 min
1          France  9/24/2021          2021  TV-MA   1 Season
2   United States  9/24/2021          2021  TV-MA   1 Season
3          Brazil  9/22/2021          2021  TV-PG     91 min
4   United States  9/24/2021          1993  TV-MA    125 min
5  United Kingdom  9/24/2021          2021  TV-14  9 Seasons
6   United States  9/24/2021          2021  PG-13    104 min
7           India   5/1/2021          2019  TV-Y7     87 min
8         Germany  9/23/2021          2021  TV-MA    127 min
9           India   5/1/2021          2013  TV-Y7     76 min

                                           listed_in
0                                      Documentaries
1  Crime TV Shows, International TV Shows, TV Act...
2                 TV Dramas, TV Horror, TV Mysteries
3                 Children & Family Movies, Comedies
4    Dramas, Independent Movies, International Movies
5                       British TV Shows, Reality TV
6                                   Comedies, Dramas
7  Children & Family Movies, Comedies, Music & Mu...
8                       Dramas, International Movies
9          Children & Family Movies, Music & Musicals
```

[14] `data.columns`

```
data.columns
```

```
Index(['show_id', 'type', 'title', 'director', 'country', 'date_added',
       'release_year', 'rating', 'duration', 'listed_in'],
      dtype='object')
```

[15]
```
data.type.unique()
```

```
array(['Movie', 'TV Show'], dtype=object)
```

[16]
```
from google.colab import drive
drive.mount('/content/drive')
```

```
Mounted at /content/drive
```

[17]
```
data.columns
```

```
Index(['show_id', 'type', 'title', 'director', 'country', 'date_added',
       'release_year', 'rating', 'duration', 'listed_in'],
      dtype='object')
```

[18]
```
data.shape
```

```
(8790, 10)
```

[19]
```
print(data.isnull().sum())
```

```
show_id         0
type            0
title           0
director        0
country         0
date_added      0
release_year    0
rating          0
duration        0
listed_in       0
dtype: int64
```

[20]
```
data.describe()
```

	release_year
count	8790.000000
mean	2014.183163
std	8.825466
min	1925.000000
25%	2013.000000
50%	2017.000000
75%	2019.000000
max	2021.000000

[21]
```
columns=list(data.columns)
print(columns)
```

```
['show_id', 'type', 'title', 'director', 'country', 'date_added', 'release_year', 'rating', 'duration', 'listed_in']
```

```python
[22] data.count()
```

	0
show_id	8790
type	8790
title	8790
director	8790
country	8790
date_added	8790
release_year	8790
rating	8790
duration	8790
listed_in	8790

dtype: int64

Start coding or generate with AI.

```python
[23] data.isnull() #here we find no null values so that we are not using it, even though when we use this replace it will automatically
```

	show_id	type	title	director	country	date_added	release_year	rating	duration	listed_in
0	False	False	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False	False
...	...	...	...	...	...	...	...	...	...	...
8785	False	False	False	False	False	False	False	False	False	False
8786	False	False	False	False	False	False	False	False	False	False
8787	False	False	False	False	False	False	False	False	False	False
8788	False	False	False	False	False	False	False	False	False	False
8789	False	False	False	False	False	False	False	False	False	False

8790 rows × 10 columns

```python
[24] print(data.isnull().sum()) #in this we have no null values are detected in data that swhy here itshows all zeros if any found nulls it will be counted as 1,2,3,....
```

```
show_id         0
type            0
title           0
director        0
country         0
date_added      0
release_year    0
rating          0
duration        0
listed_in       0
dtype: int64
```

```
[25]
```

```
[25]    #remove if founnd any null values , imean 0's
        #(data[columns[8790:10]]==0).sum()
        data [columns[8790:10]]=data[columns[8790:10]].replace(0,np.nan)
```

```
[26]    data.columns
```

```
Index(['show_id', 'type', 'title', 'director', 'country', 'date_added',
       'release_year', 'rating', 'duration', 'listed_in'],
      dtype='object')
```

[ ] Start coding or generate with AI.

```
[27]    # prompt: i need to replace the values from not given to nan

        data = data.replace('0', np.nan)
```

```
[ ]    data.columns
```

```
Index(['show_id', 'type', 'title', 'director', 'country', 'date_added',
       'release_year', 'rating', 'duration', 'listed_in'],
      dtype='object')
```

```
[29]    data.shape #there is no change in before data and after data
```

```
(8790, 10)
```

+ Code   + Text

```
[30]    data.dropna(inplace=True) #here we remove when we get not accessable values like 0's
```

```
[31]    data.dropna(inplace=True)
        data   #here if we see in Directors column there are not given names .... if we want to remove it , we can use in separatly
```

	show_id	type	title	director	country	date_added	release_year	rating	duration	listed_in
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	United States	9/25/2021	2020	PG-13	90 min	Documentaries
1	s3	TV Show	Ganglands	Julien Leclercq	France	9/24/2021	2021	TV-MA	1 Season	Crime TV Shows, International TV Shows, TV Act...
2	s6	TV Show	Midnight Mass	Mike Flanagan	United States	9/24/2021	2021	TV-MA	1 Season	TV Dramas, TV Horror, TV Mysteries
3	s14	Movie	Confessions of an Invisible Girl	Bruno Garotti	Brazil	9/22/2021	2021	TV-PG	91 min	Children & Family Movies, Comedies
4	s8	Movie	Sankofa	Haile Gerima	United States	9/24/2021	1993	TV-MA	125 min	Dramas, Independent Movies, International Movies
...	...	...	...	...	...	...	...	...	...	...
8785	s8797	TV Show	Yunus Emre	Not Given	Turkey	1/17/2017	2016	TV-PG	2 Seasons	International TV Shows, TV Dramas
8786	s8798	TV Show	Zak Storm	Not Given	United States	9/13/2018	2016	TV-Y7	3 Seasons	Kids' TV
8787	s8801	TV Show	Zindagi Gulzar Hai	Not Given	Pakistan	12/15/2016	2012	TV-PG	1 Season	International TV Shows, Romantic TV Shows, TV ...
8788	s8784	TV Show	Yoko	Not Given	Pakistan	6/23/2018	2016	TV-Y	1 Season	Kids' TV
8789	s8786	TV Show	YOM	Not Given	Pakistan	6/7/2018	2016	TV-Y7	1 Season	Kids' TV

8790 rows × 10 columns

Next steps: Generate code with data   👁 View recommended plots   New interactive sheet

```
[32]    data.columns
        #here im not deleting Directors column becuase we need to find Top10 Directors
```

```
Index(['show_id', 'type', 'title', 'director', 'country', 'date_added',
       'release_year', 'rating', 'duration', 'listed_in'],
```

```python
data=pd.read_csv('/content/netflix1.csv')
```

[12] `data.head()`

	show_id	type	title	director	country	date_added	release_year	rating	duration	listed_in
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	United States	9/25/2021	2020	PG-13	90 min	Documentaries
1	s3	TV Show	Ganglands	Julien Leclercq	France	9/24/2021	2021	TV-MA	1 Season	Crime TV Shows, International TV Shows, TV Act...
2	s6	TV Show	Midnight Mass	Mike Flanagan	United States	9/24/2021	2021	TV-MA	1 Season	TV Dramas, TV Horror, TV Mysteries
3	s14	Movie	Confessions of an Invisible Girl	Bruno Garotti	Brazil	9/22/2021	2021	TV-PG	91 min	Children & Family Movies, Comedies
4	s8	Movie	Sankofa	Haile Gerima	United States	9/24/2021	1993	TV-MA	125 min	Dramas, Independent Movies, International Movies

Next steps: Generate code with `data`   ⦿ View recommended plots   New interactive sheet

```python
print(data.head(10))
```

```
  show_id     type                             title             director  \
0      s1    Movie              Dick Johnson Is Dead      Kirsten Johnson
1      s3  TV Show                         Ganglands      Julien Leclercq
2      s6  TV Show                     Midnight Mass        Mike Flanagan
3     s14    Movie  Confessions of an Invisible Girl        Bruno Garotti
4      s8    Movie                           Sankofa         Haile Gerima
5      s9  TV Show         The Great British Baking Show      Andy Devonshire
6     s10    Movie                      The Starling       Theodore Melfi
7    s939    Movie       Motu Patlu in the Game of Zones         Suhas Kadav
8     s13    Movie                       Je Suis Karl  Christian Schwochow
9    s940    Movie           Motu Patlu in Wonderland         Suhas Kadav

          country date_added  release_year rating   duration  \
0   United States  9/25/2021          2020  PG-13     90 min
1          France  9/24/2021          2021  TV-MA   1 Season
2   United States  9/24/2021          2021  TV-MA   1 Season
3          Brazil  9/22/2021          2021  TV-PG     91 min
4   United States  9/24/2021          1993  TV-MA    125 min
5  United Kingdom  9/24/2021          2021  TV-14  9 Seasons
6   United States  9/24/2021          2021  PG-13    104 min
7           India   5/1/2021          2019   TV-Y7     87 min
8         Germany  9/23/2021          2021  TV-MA    127 min
9           India   5/1/2021          2013   TV-Y7     76 min

                                           listed_in
0                                      Documentaries
1  Crime TV Shows, International TV Shows, TV Act...
2                 TV Dramas, TV Horror, TV Mysteries
3                 Children & Family Movies, Comedies
4    Dramas, Independent Movies, International Movies
5                      British TV Shows, Reality TV
6                                  Comedies, Dramas
7  Children & Family Movies, Comedies, Music & Mu...
8                        Dramas, International Movies
9            Children & Family Movies, Music & Musicals
```

[14] `data.columns`

```
Index(['show_id', 'type', 'title', 'director', 'country', 'date_added',
       'release_year', 'rating', 'duration', 'listed_in'],
      dtype='object')
```

[15]
```python
data.type.unique()
```

```
array(['Movie', 'TV Show'], dtype=object)
```

✓ 0s   completed at 12:28 PM

```
array(['Movie', 'TV Show'], dtype=object)
```

```python
[16] from google.colab import drive
     drive.mount('/content/drive')
```

Mounted at /content/drive

```python
[17] data.columns
```

```
Index(['show_id', 'type', 'title', 'director', 'country', 'date_added',
       'release_year', 'rating', 'duration', 'listed_in'],
      dtype='object')
```

```python
[18] data.shape
```

```
(8790, 10)
```

```python
[19] print(data.isnull().sum())
```

```
show_id           0
type              0
title             0
director          0
country           0
date_added        0
release_year      0
rating            0
duration          0
listed_in         0
dtype: int64
```

```python
[20] data.describe()
```

	release_year
count	8790.000000
mean	2014.183163
std	8.825466
min	1925.000000
25%	2013.000000
50%	2017.000000
75%	2019.000000
max	2021.000000

```python
[21] columns=list(data.columns)
     print(columns)
```

```
['show_id', 'type', 'title', 'director', 'country', 'date_added', 'release_year', 'rating', 'duration', 'listed_in']
```

```python
data.count()
```

	0
show_id	8790
type	8790
title	8790
director	8790

```python
data.isnull() #here we find no null values so that we are not using it, even though when we use this replace it will automatically
```

	show_id	type	title	director	country	date_added	release_year	rating	duration	listed_in
0	False	False	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False	False
...	...	...	...	...	...	...	...	...	...	...
8785	False	False	False	False	False	False	False	False	False	False
8786	False	False	False	False	False	False	False	False	False	False
8787	False	False	False	False	False	False	False	False	False	False
8788	False	False	False	False	False	False	False	False	False	False
8789	False	False	False	False	False	False	False	False	False	False

8790 rows × 10 columns

```python
[24] print(data.isnull().sum()) #in this we have no null values are detected in data that swhy here itshows all zeros if any found nulls it will be counted as 1,2,3,....
```

```
show_id          0
type             0
title            0
director         0
country          0
date_added       0
release_year     0
rating           0
duration         0
listed_in        0
dtype: int64
```

```python
[25] #remove if founnd any null values , imean 0's
     #(data[columns[8790:10]]==0).sum()
     data [columns[8790:10]]=data[columns[8790:10]].replace(0,np.nan)
```

```python
[26] data.columns
```

```
Index(['show_id', 'type', 'title', 'director', 'country', 'date_added',
       'release_year', 'rating', 'duration', 'listed_in'],
      dtype='object')
```

```python
[27] # prompt: i need to replace the values from not given to nan

     data = data.replace('0', np.nan)
```

```python
[28] data.columns
```

```
Index(['show_id', 'type', 'title', 'director', 'country', 'date_added',
       'release_year', 'rating', 'duration', 'listed_in'],
      dtype='object')
```

```
(8790, 10)
```

```
[30] data.dropna(inplace=True) #here we remove when we get not accessable values like 0's
```

```
[31] data.dropna(inplace=True)
     data    #here if we see in Directors column there are not given names .... if we want to remove it , we can use in separatly
```

	show_id	type	title	director	country	date_added	release_year	rating	duration	listed_in
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	United States	9/25/2021	2020	PG-13	90 min	Documentaries
1	s3	TV Show	Ganglands	Julien Leclercq	France	9/24/2021	2021	TV-MA	1 Season	Crime TV Shows, International TV Shows, TV Act...
2	s6	TV Show	Midnight Mass	Mike Flanagan	United States	9/24/2021	2021	TV-MA	1 Season	TV Dramas, TV Horror, TV Mysteries
3	s14	Movie	Confessions of an Invisible Girl	Bruno Garotti	Brazil	9/22/2021	2021	TV-PG	91 min	Children & Family Movies, Comedies
4	s8	Movie	Sankofa	Haile Gerima	United States	9/24/2021	1993	TV-MA	125 min	Dramas, Independent Movies, International Movies
...	...	...	...	...	...	...	...	...	...	...
8785	s8797	TV Show	Yunus Emre	Not Given	Turkey	1/17/2017	2016	TV-PG	2 Seasons	International TV Shows, TV Dramas
8786	s8798	TV Show	Zak Storm	Not Given	United States	9/13/2018	2016	TV-Y7	3 Seasons	Kids' TV
8787	s8801	TV Show	Zindagi Gulzar Hai	Not Given	Pakistan	12/15/2016	2012	TV-PG	1 Season	International TV Shows, Romantic TV Shows, TV ...
8788	s8784	TV Show	Yoko	Not Given	Pakistan	6/23/2018	2016	TV-Y	1 Season	Kids' TV
8789	s8786	TV Show	YOM	Not Given	Pakistan	6/7/2018	2016	TV-Y7	1 Season	Kids' TV

8790 rows × 10 columns

Next steps: Generate code with `data`   ◑ View recommended plots   New interactive sheet

```
[32] data.columns
     #here im not deleting Directors column becuase we need to find Top10 Directors
```

```
Index(['show_id', 'type', 'title', 'director', 'country', 'date_added',
       'release_year', 'rating', 'duration', 'listed_in'],
      dtype='object')
```

```
data.drop_duplicates(inplace=True)
print(data.drop_duplicates)
```

```
<bound method DataFrame.drop_duplicates of        show_id     type                             title          director  \
0           s1    Movie              Dick Johnson Is Dead   Kirsten Johnson
1           s3  TV Show                         Ganglands   Julien Leclercq
2           s6  TV Show                     Midnight Mass     Mike Flanagan
3          s14    Movie  Confessions of an Invisible Girl     Bruno Garotti
4           s8    Movie                           Sankofa      Haile Gerima
...        ...      ...                               ...               ...
8785     s8797  TV Show                        Yunus Emre         Not Given
8786     s8798  TV Show                         Zak Storm         Not Given
8787     s8801  TV Show                Zindagi Gulzar Hai         Not Given
8788     s8784  TV Show                              Yoko         Not Given
8789     s8786  TV Show                               YOM         Not Given

            country date_added  release_year rating    duration  \
0     United States  9/25/2021          2020  PG-13      90 min
1            France  9/24/2021          2021  TV-MA    1 Season
2     United States  9/24/2021          2021  TV-MA    1 Season
3            Brazil  9/22/2021          2021  TV-PG      91 min
4     United States  9/24/2021          1993  TV-MA     125 min
...             ...        ...           ...    ...         ...
8785         Turkey  1/17/2017          2016  TV-PG   2 Seasons
8786  United States  9/13/2018          2016  TV-Y7   3 Seasons
8787       Pakistan 12/15/2016          2012  TV-PG    1 Season
8788       Pakistan  6/23/2018          2016   TV-Y    1 Season
```

```
                            [8790 rows x 10 columns]>
```

```
[34]  data.dropna(subset=['director', 'country'],inplace=True)
      data
```

	show_id	type	title	director	country	date_added	release_year	rating	duration	listed_in
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	United States	9/25/2021	2020	PG-13	90 min	Documentaries
1	s3	TV Show	Ganglands	Julien Leclercq	France	9/24/2021	2021	TV-MA	1 Season	Crime TV Shows, International TV Shows, TV Act...
2	s6	TV Show	Midnight Mass	Mike Flanagan	United States	9/24/2021	2021	TV-MA	1 Season	TV Dramas, TV Horror, TV Mysteries
3	s14	Movie	Confessions of an Invisible Girl	Bruno Garotti	Brazil	9/22/2021	2021	TV-PG	91 min	Children & Family Movies, Comedies
4	s8	Movie	Sankofa	Haile Gerima	United States	9/24/2021	1993	TV-MA	125 min	Dramas, Independent Movies, International Movies
...	...	...	...	...	...	...	...	...	...	...
8785	s8797	TV Show	Yunus Emre	Not Given	Turkey	1/17/2017	2016	TV-PG	2 Seasons	International TV Shows, TV Dramas
8786	s8798	TV Show	Zak Storm	Not Given	United States	9/13/2018	2016	TV-Y7	3 Seasons	Kids' TV
8787	s8801	TV Show	Zindagi Gulzar Hai	Not Given	Pakistan	12/15/2016	2012	TV-PG	1 Season	International TV Shows, Romantic TV Shows, TV ...
8788	s8784	TV Show	Yoko	Not Given	Pakistan	6/23/2018	2016	TV-Y	1 Season	Kids' TV
8789	s8786	TV Show	YOM	Not Given	Pakistan	6/7/2018	2016	TV-Y7	1 Season	Kids' TV

8790 rows × 10 columns

Next steps: Generate code with `data`   View recommended plots   New interactive sheet

```
[35]  data.shape
```

```
(8790, 10)
```

```
[36]  data['date_added'] = pd.to_datetime(data['date_added'])
```

```
[37]  print(data.dtypes)
```

```
show_id               object
type                  object
title                 object
director              object
country               object
date_added    datetime64[ns]
release_year           int64
rating                object
duration              object
listed_in             object
dtype: object
```

```
#After Cleaning of data we are going to E-Data Analysis
# ! 1.content Type Distribution (Movies Vs Tv Shows)

type_counts = data['type'].value_counts()
print(type_counts)
```
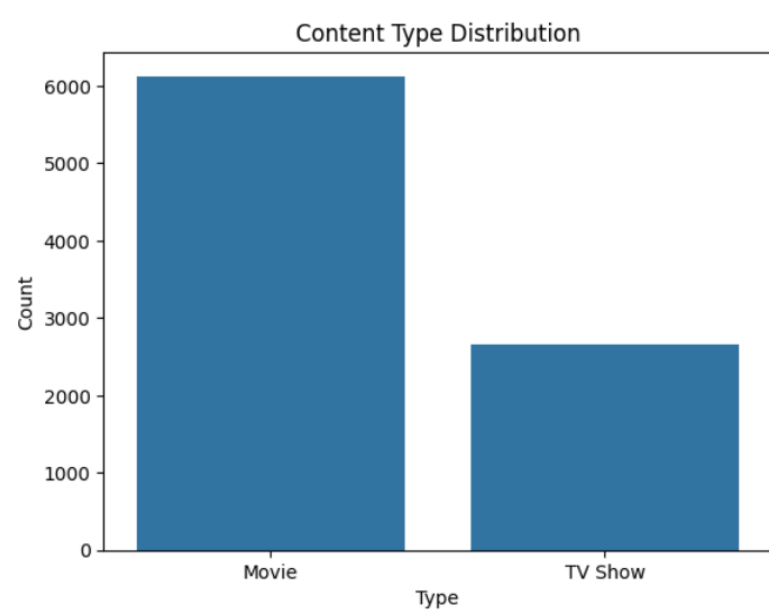
```
type
Movie      6126
TV Show    2664
Name: count, dtype: int64
```
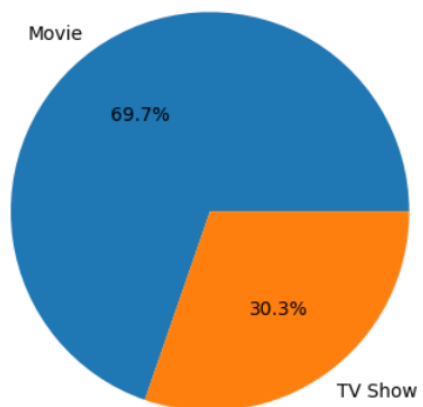
```
[39] #Now we unnder go plot the data  we use matplot library

     sns.barplot(x=type_counts.index, y=type_counts.values)
     plt.title('Content Type Distribution')
     plt.xlabel('Type')
     plt.ylabel('Count')
     plt.show()


     plt.pie(type_counts.values, labels=type_counts.index, autopct='%1.1f%%')
     plt.title('Content Type Distribution')
     plt.show()
```


Content Type Distribution

Content Type Distribution

Here we have X-axis as Type and Y-axis have Count and Movies are watching are 6000 where as Tv-Shows comes under 2500 Content Type Disribution.

```
[ ]  Start coding or generate with AI.
```

```
[40] #2. MOst Common Genres
```

```
[41] data [ 'genres']=data['listed_in'].apply(lambda x:x.split(','))
     all_genres=sum(data['genres'],[])
     genre_counts=pd.Series(all_genres).value_counts().head(10)
```
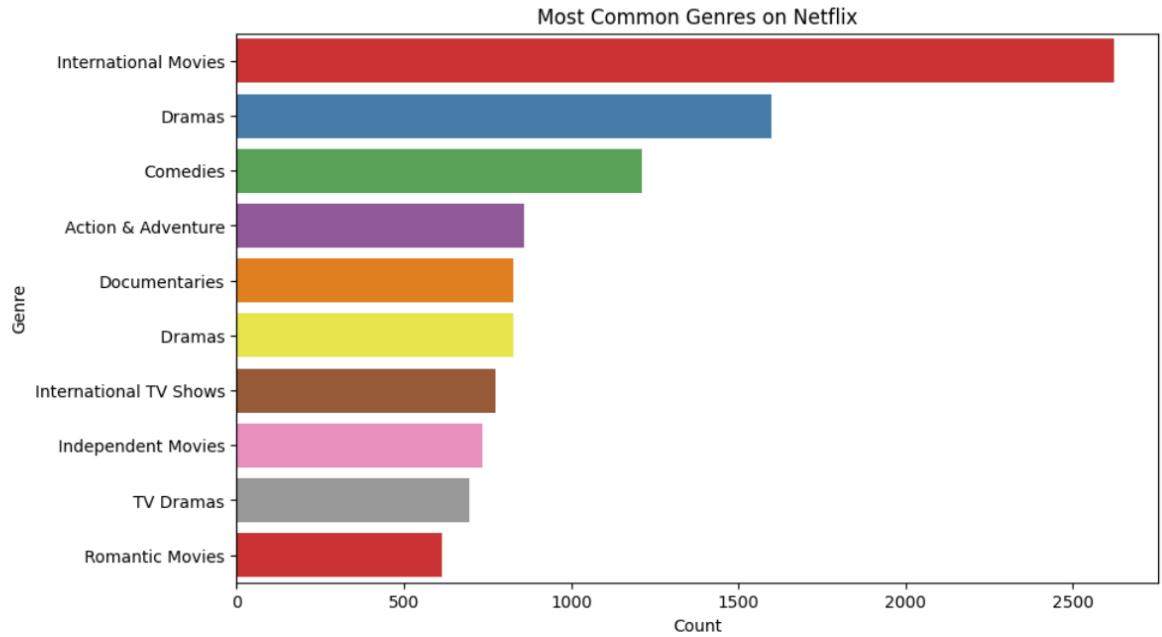
```
[42] all_genres = sum (data['genres'], [])
     genres_counts=pd.Series(all_genres).value_counts().head(10)
```

```
plt.figure(figsize=(10,6))
sns.barplot(x=genre_counts.values,y=genre_counts.index,palette='Set1')
plt.title('Most Common Genres on Netflix')
plt.xlabel('Count')
plt.ylabel('Genre')
plt.show()
data.genres
```

```
<ipython-input-43-eae1bd32fefa>:2: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `y` variable to `hue` and set `legend=False` for the same effect.

  sns.barplot(x=genre_counts.values,y=genre_counts.index,palette='Set1')
```


	genres
0	[Documentaries]
1	[Crime TV Shows, International TV Shows, TV ...
2	[TV Dramas, TV Horror, TV Mysteries]
3	[Children & Family Movies, Comedies]

```
+ Code    + Text
```

	genres
0	[Documentaries]
1	[Crime TV Shows, International TV Shows, TV ...
2	[TV Dramas, TV Horror, TV Mysteries]
3	[Children & Family Movies, Comedies]
4	[Dramas, Independent Movies, International M...
...	...
8785	[International TV Shows, TV Dramas]
8786	[Kids' TV]
8787	[International TV Shows, Romantic TV Shows, ...
8788	[Kids' TV]
8789	[Kids' TV]

8790 rows × 1 columns

dtype: object

```python
[44]  #3. Content Added Over Time
      # Extract year and month from 'date_added'
```

```python
[45]  data.columns
```

```
Index(['show_id', 'type', 'title', 'director', 'country', 'date_added',
       'release_year', 'rating', 'duration', 'listed_in', 'genres'],
      dtype='object')
```

```python
[46]  data['year_added']=data['date_added'].dt.year
      data['month_added']=data['date_added'].dt.month
```

```python
[47]  plt.figure(figsize=(12,6)) #we are using Default statement
      sns.countplot(x='year_added',data=data,palette='coolwarm')
      plt.title('Content Added Over Time')
      plt.xlabel('Year')
      plt.ylabel('Count')
      plt.xticks(rotation=90) #rotation changes on X axis itself .....
```
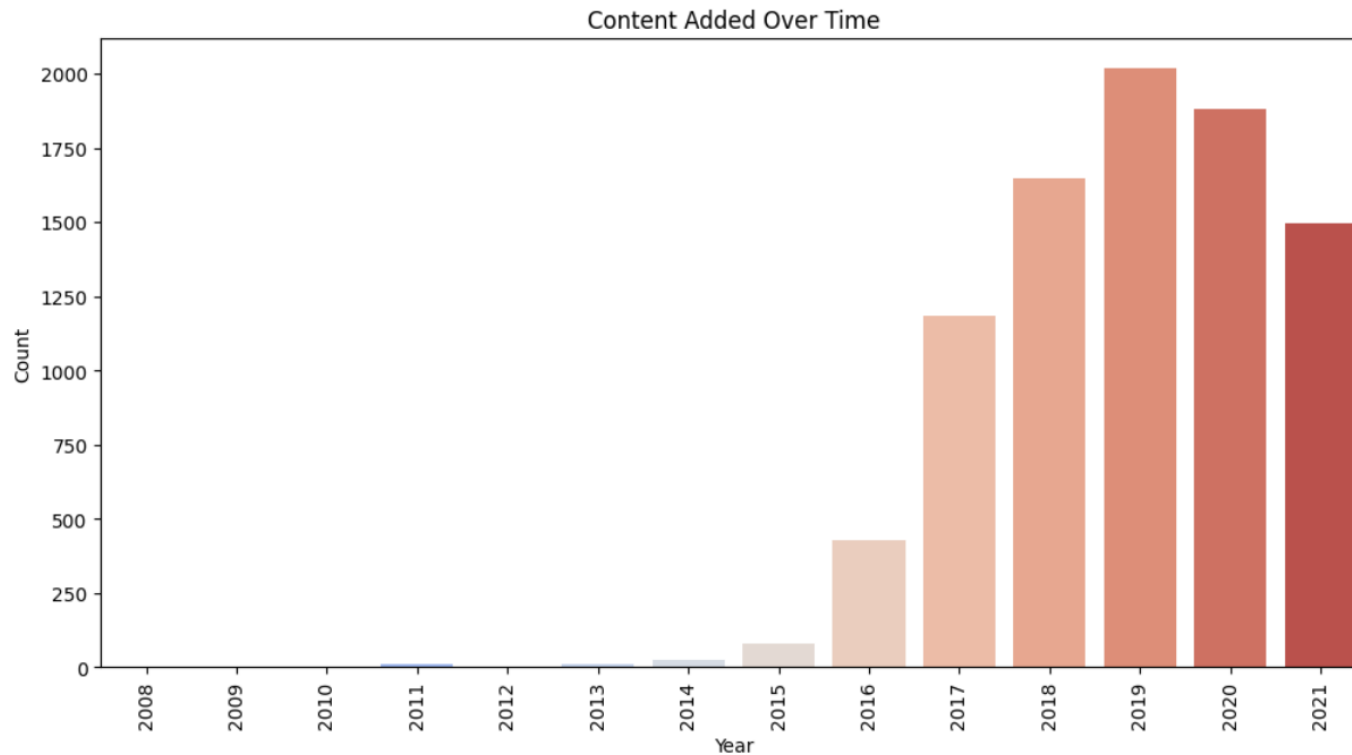
```
<ipython-input-47-7ed27e383996>:2: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same effect.

  sns.countplot(x='year_added',data=data,palette='coolwarm')
([0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13],
 [Text(0, 0, '2008'),
  Text(1, 0, '2009'),
  Text(2, 0, '2010'),
  Text(3, 0, '2011'),
  Text(4, 0, '2012'),
  Text(5, 0, '2013'),
  Text(6, 0, '2014'),
  Text(7, 0, '2015'),
  Text(8, 0, '2016'),
  Text(9, 0, '2017'),
  Text(10, 0, '2018'),
  Text(11, 0, '2019'),
  Text(12, 0, '2020'),
  Text(13, 0, '2021')])
```

Content Added Over Time

```
plt.xlabel('Year')
plt.ylabel('Count')
plt.xticks(rotation=90) #rotation changes on X axis itself .....
```

```
<ipython-input-47-7ed27e383996>:2: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same effect.

  sns.countplot(x='year_added',data=data,palette='coolwarm')
([0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13],
 [Text(0, 0, '2008'),
  Text(1, 0, '2009'),
  Text(2, 0, '2010'),
  Text(3, 0, '2011'),
  Text(4, 0, '2012'),
  Text(5, 0, '2013'),
  Text(6, 0, '2014'),
  Text(7, 0, '2015'),
  Text(8, 0, '2016'),
  Text(9, 0, '2017'),
  Text(10, 0, '2018'),
  Text(11, 0, '2019'),
  Text(12, 0, '2020'),
  Text(13, 0, '2021')])
```

Content Added Over Time

by analysing all the years Movies and Tv shows were more increases from 2016 to 2021 when compared to 2008-2015 and both genres counts also increases and then decreases from 2020.

[48] #4. Top 10 Directors with the Most Titles

```
[49] top_directors = data['director'].value_counts().head(10)
```

```
[50]   # 4. Top 10 Directors with the Most Titles
       top_directors = data['director'].value_counts().head(10)

       # Remove rows where 'director' is 'Not Given'
       top_directors = top_directors[top_directors.index != 'Not Given']

       plt.figure(figsize=(10,6))
       sns.barplot(x=top_directors.values,y=top_directors.index,palette='viridis')
       plt.title('Top 10 Directors with the Most Titles')
       plt.xlabel('Count')
       plt.ylabel('Director')
       plt.show()
```
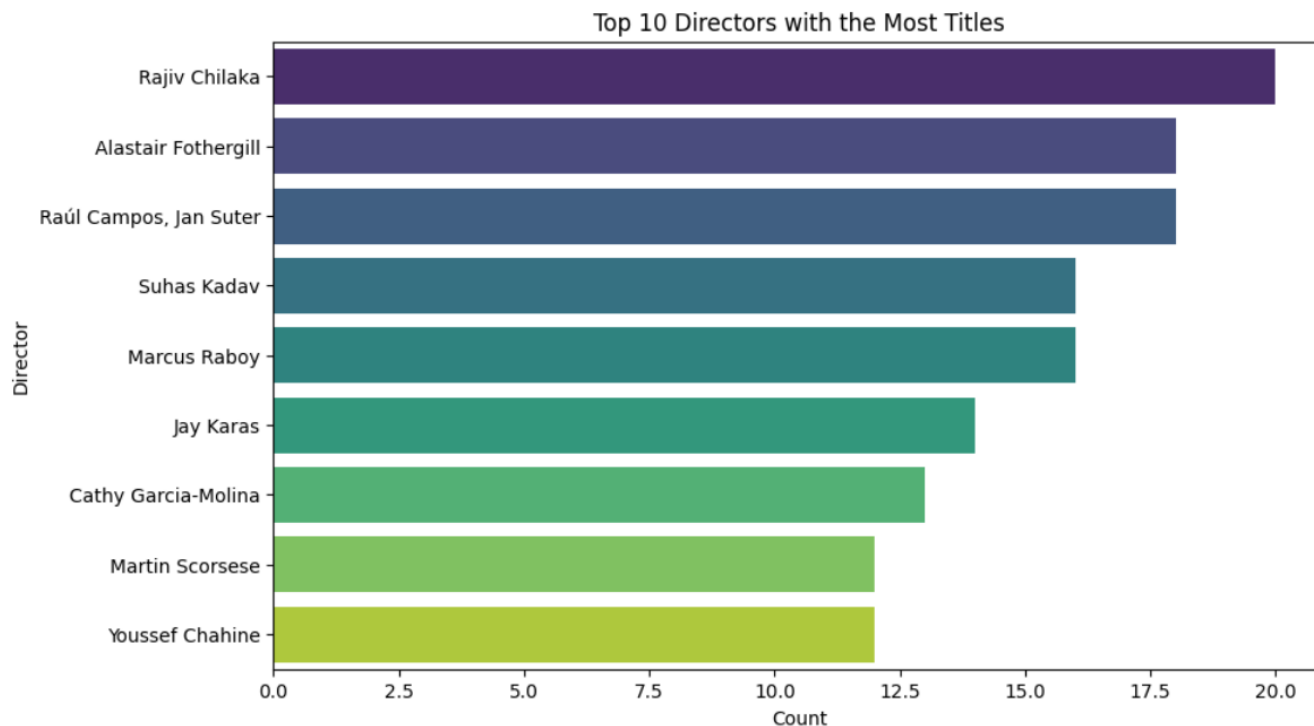
<ipython-input-50-d9963413962e>:8: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `y` variable to `hue` and set `legend=False` for the same effect.

    sns.barplot(x=top_directors.values,y=top_directors.index,palette='viridis')
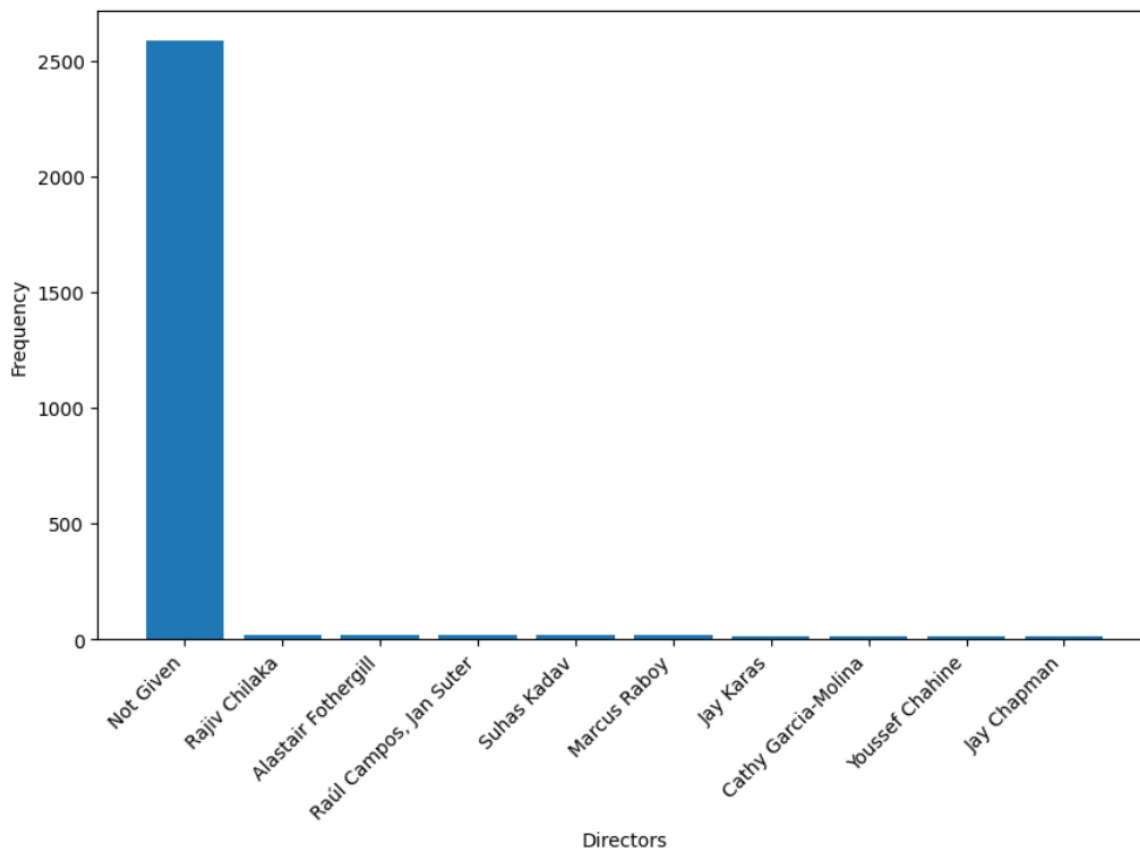

Top 10 Directors with the Most Titles

```
top_ten_directors=data['director'].value_counts().reset_index().sort_values(by='count', ascending=False)[:10]
plt.figure(figsize=(10, 6))
plt.bar(top_ten_directors['director'], top_ten_directors['count'])
plt.xticks(rotation=45, ha='right')
plt.xlabel("Directors")
plt.ylabel("Frequency")
plt.suptitle("Top 10 countries with most content on Netflix")
plt.show()
```

Top 10 countries with most content on Netflix

Top 10 countries with most content on Netflix

[52]
```python
# Remove rows where 'director' is "Not Given"
data = data[data['director'] != 'Not Given']
```

```python
top_ten_directors=data['director'].value_counts().reset_index().sort_values(by='count', ascending=False)[:10]
plt.figure(figsize=(10, 6))
plt.bar(top_ten_directors['director'], top_ten_directors['count'])
plt.xticks(rotation=45, ha='right')
plt.xlabel("Directors")
plt.ylabel("Frequency")
plt.title("Top 10 countries with most content on Netflix")
plt.show()

#
```

Top 10 countries with most content on Netflix

[52]
```python
# Remove rows where 'director' is "Not Given"
data = data[data['director'] != 'Not Given']
```

```python
top_ten_directors=data['director'].value_counts().reset_index().sort_values(by='count', ascending=False)[:10]
plt.figure(figsize=(10, 6))
plt.bar(top_ten_directors['director'], top_ten_directors['count'])
plt.xticks(rotation=45, ha='right')
plt.xlabel("Directors")
plt.ylabel("Frequency")
plt.title("Top 10 countries with most content on Netflix")
plt.show()

#
```
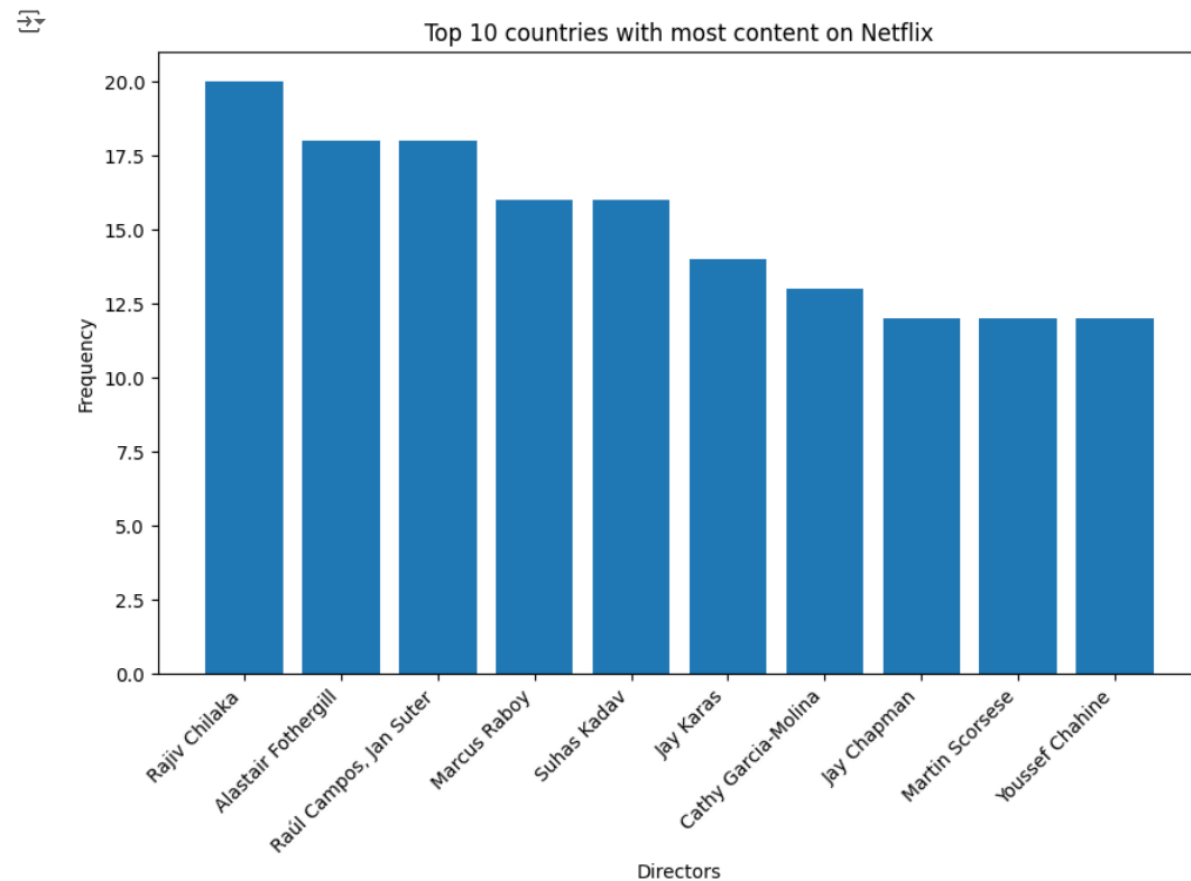

Top 10 countries with most content on Netflix
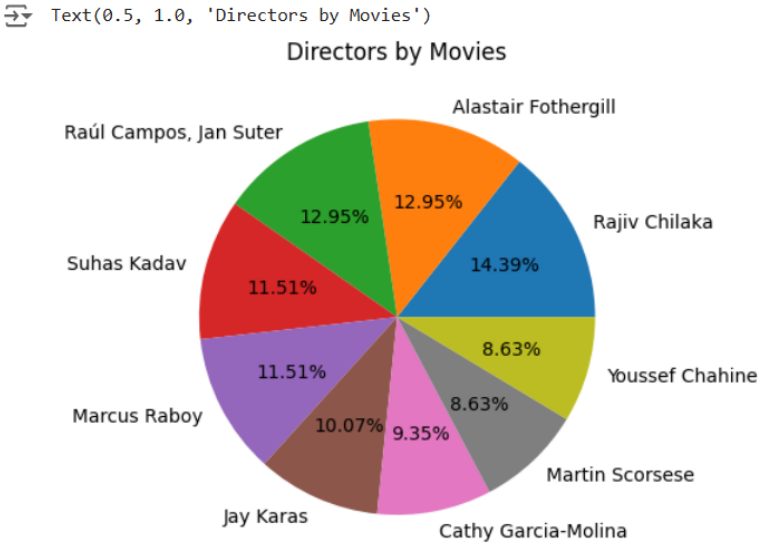
[54]
```python
data.shape
```

(6202, 13)

[55]
```python
#After importing Library
data.columns
```

```
Index(['show_id', 'type', 'title', 'director', 'country', 'date_added',
       'release_year', 'rating', 'duration', 'listed_in', 'genres',
       'year_added', 'month_added'],
      dtype='object')
```

```
[84]  # Assuming 'data' DataFrame is already created and processed as in the provided code.

      # Word cloud for movie titles
      movie_titles = data[data['type'] == 'Movie']['title']
      wordcloud = WordCloud(width=800, height=400, background_color='black').generate(' '.join(movie_titles))

      # Plot the word cloud
      plt.figure(figsize=(10, 6))
      plt.imshow(wordcloud, interpolation='bilinear')
      plt.axis('off')
      plt.title('Word Cloud of Movie Titles') # Added title for better visualization
      plt.show()
```

Word Cloud of Movie Titles


```
data['type'].value_counts()
```

	count
type	
Movie	5953
TV Show	249

dtype: int64

```
[59]  data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 6202 entries, 0 to 6569
Data columns (total 13 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   show_id       6202 non-null   object
 1   type          6202 non-null   object
 2   title         6202 non-null   object
 3   director      6202 non-null   object
 4   country       6202 non-null   object
 5   date_added    6202 non-null   datetime64[ns]
 6   release_year  6202 non-null   int64
 7   rating        6202 non-null   object
 8   duration      6202 non-null   object
 9   listed_in     6202 non-null   object
 10  genres        6202 non-null   object
```

```
[85]  data.to_csv('cleaned_netflix_data1.csv', index=False)
      # new Data is Craeated after Cleaning all the unkonwn values
```

```
[69]  plt.pie(labels=top_directors.index,x=top_directors.values,autopct='%1.2f%%',)
      plt.title('Directors by Movies')
```

Text(0.5, 1.0, 'Directors by Movies')
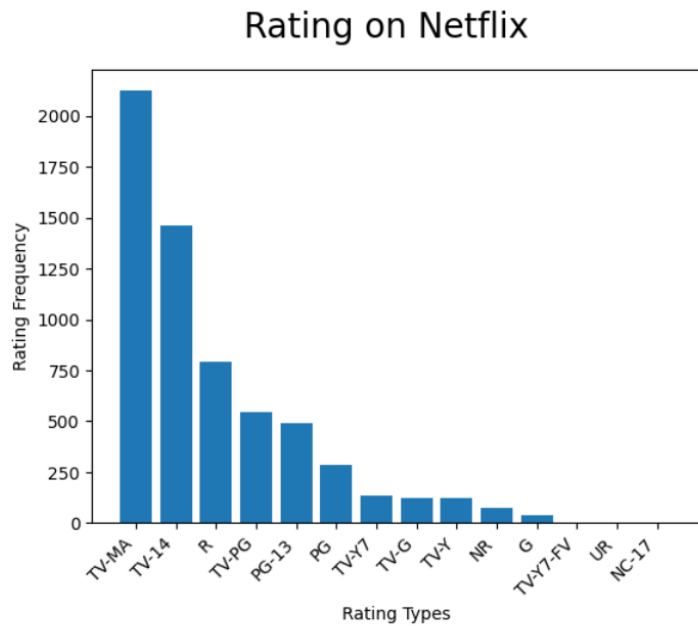
Directors by Movies


```
data['rating'].value_counts()
```

	count
rating	
TV-MA	2123
TV-14	1460
R	795
TV-PG	547
PG-13	489
PG	286
TV-Y7	133
TV-G	123
TV-Y	121
NR	75
G	41
TV-Y7-FV	4
UR	3
NC-17	2

dtype: int64

```
[70] ratings=data['rating'].value_counts().reset_index().sort_values(by='count',ascending=False)
     plt.bar(ratings['rating'], ratings['count'])
     plt.xticks(rotation=45, ha='right')
     plt.xlabel("Rating Types")
     plt.ylabel("Rating Frequency")
     plt.suptitle('Rating on Netflix', fontsize=20)
```

Text(0.5, 0.98, 'Rating on Netflix')


```
plt.pie(ratings['count'][:8], labels=ratings['rating'][:8], autopct='%.0f%%')
plt.suptitle('Rating on Netflix', fontsize=20)
```

Text(0.5, 0.98, 'Rating on Netflix')